STATUS OF INFORMATION COLLECTION REQUEST (ICR) SURVEY DATABASE

The purpose of this note is to update you on the status of the survey database and EPA's expectations for when a new version of the database will be released. Version 1 (v1) of the Survey database was released on October 15. This database contains the responses to the ICR surveys EPA sent to owners of incinerators, boilers, and process heaters burning non-fossil materials. The database was created by a scanning company who scanned in the check box responses and keypunched number and text responses. An alert released with Survey v1 discussed several quality issues associated with v1, and indicated that a revised version of the database would be released when quality assurance (QA) is completed and the problems are corrected.

As described in the October 15 alert, the main problems with Survey v1 were: some data, check marks, and comments that are clearly marked on the survey responses were not in the database; some check marks were misplaced (i.e. the respondent checked one box, but in the database it appeared they had checked another); there were some typographical errors in data and comment entry; and some description and comment fields were truncated. The scanner was notified of these problems and asked to fix programming errors that led to the missed and misplaced check marks, and to perform a 100% QA on the hand-entered portions of the database.
The company provided a revised database on November 25. EPA and EPA's contractor, ERG, reviewed this database and met with the scanner on December 8. While the database received on November 25 showed significant improvement, several quality problems remain, and EPA has decided not to release this version of the database until these remaining problems are corrected.

Based on ERG's comparison of 300 sample survey forms with the database, the latest version of the database still has problems of misplaced check marks. The scanning company had a switch in programmers due to illness, and was not able to correct all sections of the program as of November 25, but they are confident they can complete the programming corrections so that all check marks are accurately recorded in the database. We also found several remaining errors in the keypunched portions of the survey. Many of the errors found were of a type that would not be easily recognized by a keypunch operator at the scanning company comparing a survey form with the database entry. For example, some responses expressed numbers with decimal places or as fractions rather than percentages, and the keypunch operators did not correctly interpret these. Some responses had answers written in a margin or between lines, and the keypunch operators did not correctly interpret were to place the information provided. When respondents typed in alternative units of measure in the capacity question, the number was recorded, but the units of measure were frequently missed by the keypunch operators. We have concluded that due to the nature of the remaining errors, it will be more effective for technical staff rather than clerical keypunch operators to perform additional QA and correct problems located. EPA has developed a specific plan for QA of number and text fields, and this plan is being implemented by ERG.

We expect to receive a revised version of the database with the check marks corrected from the scanning company by January 9. ERG will implement a series of computer queries on that database to identify outliers, inconsistent responses, missing units of measure, codes that are not on the code lists in the instructions, etc., and compare suspect data with the original surveys. For

important questions where we have discovered a significant error rate in keypunched numerical or text entries, ERG will check all of the entries in the database against the original surveys, and correct problems found. This work will occur during January and early February. After QA and correction, the database will be manipulated into the proper format (the same format as v1). We plan to release the revised database (Survey v2) in mid to late February.

We know the survey database is very important to the Work Groups, and we are working hard to get it done as quickly as possible. We also realize the great importance of having the database accurately portray the information contained in the survey responses. We appreciate your patience as we work to correct the problems in the survey database. If any additional problems arise or the survey database status changes, we will advise everyone as soon as possible.